Crystal Structure of the Worm NitFhit Rosetta Stone Protein Reveals a Nit Tetramer Binding Two Fhit Dimers

H.C. Pace and C. Brenner

Structural Biology & Bioinformatics Program, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA 19107

Loss of the *FHIT* gene leads to defects in cell death and is a frequent, early event in cancer [1]. The additional players in the Fhit cell death pathway are largely unknown. Department of Energy-sponsored research at the University of California at Los Angeles made the observation that proteins that are found physically fused in one form of life but separated in other forms of life function in the same biological pathway[2]. The value of the fusion event to predict a common pathway is strongest when the separate proteins have similar gene expression patterns and have a similar pattern of being present or absent in other genomes[3]. When these criteria are met, the fusion event between the two polypeptides is said to be a "Rosetta Stone" that decodes a previously hidden functional link between two proteins unrelated by sequence.

Nit and Fhit are fused in the common laboratory invertebrates *D. melanogaster* and *C. elegans* but are encoded as separate polypeptides in mammals, fish and fungi[4,5]. In the mouse, *Nit1* and *Fhit* have similar gene expression patterns in seven of eight tissues examined[4]. No form of life has been found that has a Fhit homolog that lacks a Nit homolog[5]. Thus, NitFhit fulfills the criteria for probable functional significance of a "Rosetta Stone" protein[2,3]. Further, because the worm *C. elegans* is a leading system in which to study cell death, we were eager to explore the structure and function of NitFhit.

Previous biochemical and crystallographic characterization of Fhit showed that it is dimeric with a dimer molecular weight of 35 kDa[6-8]. Because the NitFhit monomer has a predicted size of 50kDa[4], it was expected that the NitFhit molecule would weigh 100 kDa. However, analytical ultracentrifugation showed that NitFhit weighed 200 kDa, consistent with the idea that NitFhit is a tetramer[5]. Indeed, the NitFhit crystal form was in space group I222, which consists of a tetramer with 222 point symmetry located at the vertexes and the center of each unit cell. Sixty-five percent of this structure was expected to be novel Nit protein, 35% was expected to be Fhit. Further, the structure needed to explain how Nit and Fhit fit together and how NitFhit is a tetramer while Fhit is a dimer.

Selenomethionine-modified NitFhit protein was crystallized and data were collected at four wavelengths, but the data did not yield interpretable phases. Frozen crystals were thawed into compounds that produce ethylmercurial adducts on reactive cysteines. The resulting crystals diffracted more strongly and to much higher resolution than native crystals and also exhibited a change in space group from I222 to P21212. Diffraction data were collected at beamline X8C at the NSLS at the wavelength of the derivatized crystal's absorption edge and at one remote wavelength. These data were sufficient to phase the molecule by solvent flattened two-wavelength anomalous diffraction[5].

Figure 1 shows the crystal structure of the NitFhit tetramer[5]. The left panel is a cartoon representation of the tetramer while the right panel shows a ribbon diagram. NitFhit is a tetramer because Nit is a tetramer. The N-terminal Nit domain assembles into a compact tetramer containing five helices, numbered NH1 through NH5, and twelve β strands, numbered NS1 through NS12. The compact portion of the Nit domain is termed an α-β-β-α sandwich structure and Nit is the first example of this particular α-β-β-α sandwich protein fold. C-terminal to NS12, there is a final strand specified by the Nit-homologous peptide sequence, which is followed by the familiar seven strand plus two-helix fold of Fhit. The four protomers of the tetramer are color-coded green, blue, red and yellow and the Nit-homologous segments are in bold colors while the Fhit-homologous segments are in pastel. It is of interest that the C-terminal Nit strand is devoted to interacting with Fhit domains such that the NitFhit tetramer consists of a Nit tetramer binding two Fhit dimers.

NitFhit, the first Rosetta Stone protein with a solved three-dimensional structure, reveals how Nit is con-

served to interact with Fhit, implicating Nit in the Fhit death pathway. In addition, Nit contains an apparent enzyme active site that may prove to be a drug target for elimination of Fhit- tumors.

Work was supported by a grant from the National Cancer Institute (P01CA77738) to Kay Huebner, Carlo M. Croce and Charles Brenner.

References

1. L.Y.Y. Fong, V. Fidanza, N. Zanesi, L.F. Lock, L.D. Siracusa, R. Mancini, Z. Siprashvili, M. Ottey, S.E. Martin, R. Dolsky, et al.: "Muir-Torre-like syndrome in *FHIT* deficient mice," Proc. Natl. Acad. Sci. USA **97**, p. 4742-4747, 2000.
2. E.M. Marcotte, M. Pellegrini, H.L. Ng, D.W. Rice, T.O. Yeates, D. Eisenberg: "Detecting protein function and protein-protein interactions from genome sequences," Science **285**, p. 751-753, 1999.

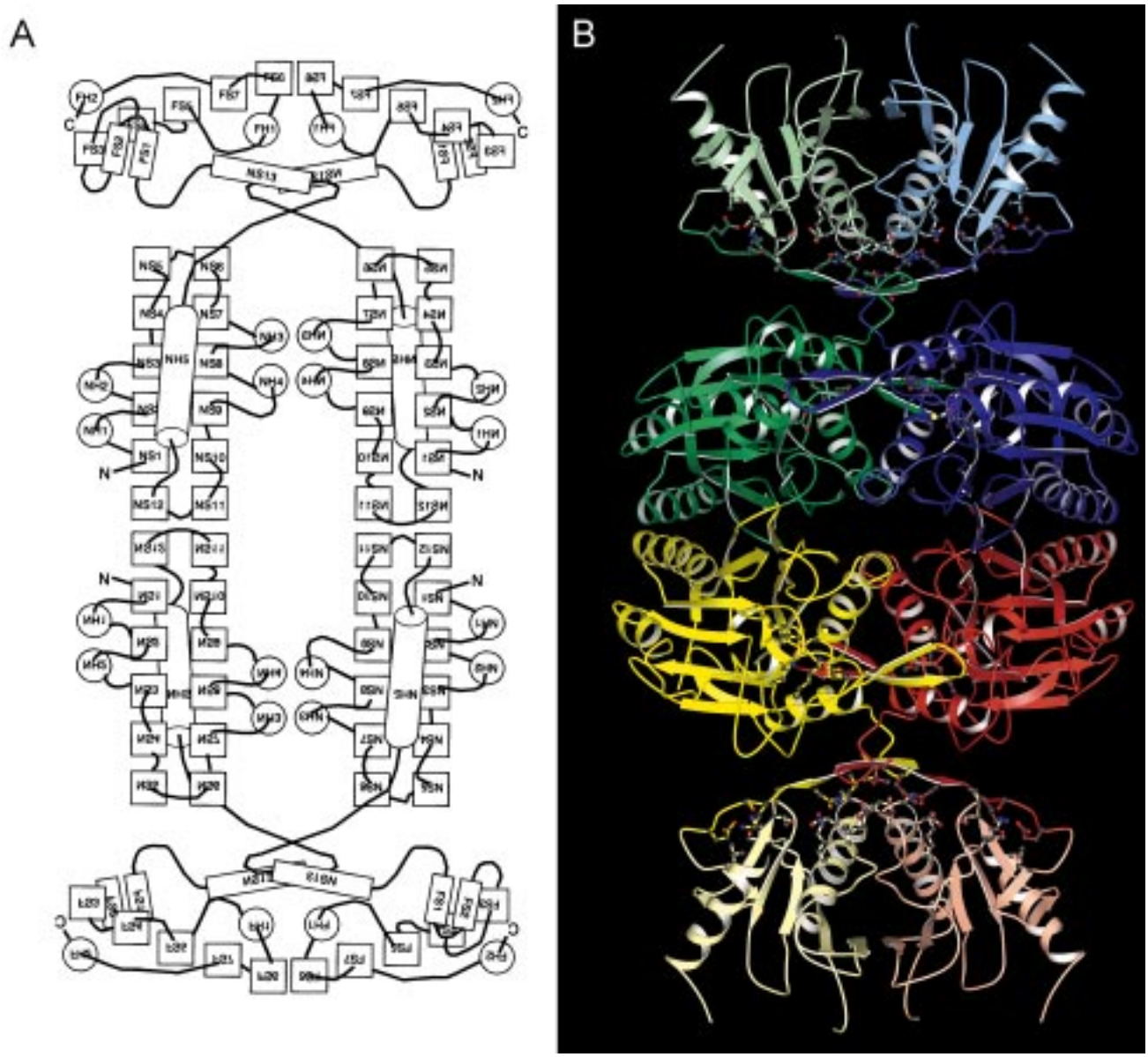

Figure 1. Crystal structure of the NitFhit tetramer. A) Cartoon representation of the NitFhit tetramer. Beta strands are represented as rectangular solids and alpha helices are represented as cylinders. Thus, the squares and circles are elements of the structure seen end-on. NitFhit monomers consist of a thirteen-strand, five helix Nit segment and a seven-strand, two helix Fhit segment. The tetramer consists of two-fold rotational symmetry about X, Y and Z. B) Ribbon representation of the NitFhit tetramer. Monomers are identified in four different colors. Nit segments are in bold. Fhit segments are in pastel.

3. E. Marcotte, M. Pellegrini, M. Thompson, T. Yeates, D. Eisenberg: "A combined algorithm for genome-wide prediction of protein function," Nature **402**, p. 83-86, 1999.
4. Y. Pekarsky, M. Campiglio, Z. Siprashvili, T. Druck, Y. Sedkov, S. Tillib, A. Draganescu, P. Wermuth, J.H. Rothman, K. Huebner, et al.: "Nitrilase and Fhit homologs are encoded as fusion proteins in Drosophila melanogaster and Caenorhabditis elegans," Proc. Natl. Acad. Sci. USA **95**, p. 8744-8749, 1998.
5. H.C. Pace, S.C. Hodawadekar, A. Draganescu, J. Huang, P. Bieganowski, Y. Pekarsky, C.M. Croce, C. Brenner: "Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers," Curr Biol **10**, p. 907-917, 2000.
6. C.D. Lima, K.L. Damico, I. Naday, G. Rosenbaum, E.M. Westbrook, W.A. Hendrickson: "MAD analysis of Fhit, a putative human tumor suppressor from the HIT protein family," Structure **5**, p. 763-774, 1997.
7. C. Brenner, H.C. Pace, P.N. Garrison, A.K. Robinson, A. Rosler, X.H. Liu, G.M. Blackburn, C.M. Croce, K. Huebner, L.D. Barnes: "Purification and crystallization of complexes modeling the active state of the fragile histidine triad protein," Protein Eng. **10**, p. 1461-1463, 1997.
8. H.C. Pace, P.N. Garrison, A.K. Robinson, L.D. Barnes, A. Draganescu, A. Rosler, G.M. Blackburn, Z. Siprashvili, C.M. Croce, K. Huebner, et al.: "Genetic, biochemical, and crystallographic characterization of Fhit- substrate complexes as the active signaling form of Fhit," Proc. Natl. Acad. Sci. USA **95**, p. 5484-5489, 1998.

Association Kinetics of the Integration Host Factor-DNA Complex by X-ray Hydroxyl Radical Footprinting

G. M. Dhavan[a], D. M. Crothers[b], M. R. Chance[a], and M. Brenowitz[a]
[a] *Center for Synchrotron Biosciences, Albert Einstein College of Medicine, Bronx, NY 10461*
[b] *Department of Chemistry, Yale University, New Haven, CT 06511*

Integration Host Factor is a sequence specific DNA binding protein that bends DNA by a considerable angle (1). It recognizes several sites in the *Escherichia coli* and lambda phage genomes (2) and plays an essential role in several biological processes, including site-specific recombination of the lambda phage, transcription, and DNA replication (3, 4, 5, 6). The crystal structure of IHF bound to one of its sites from the lambda phage genome shows that the DNA is wrapped around the protein in a U shaped turn (7).

IHF's ability to bend lambda phage DNA makes it an important architectural protein in the formation of a nucleoprotein complex that precedes the integration of the phage into the bacterial genome (8). Since IHF binding is an essential step in site-specific recombination, the kinetics of IHF-DNA complex formation affect the overall rate of the recombination process. In our work we have determined the site-specific association kinetics of IHF binding to one of its sites from the lambda phage genome, the H′ site.

Time-resolved hydroxyl radical footprinting conducted at beamline X28C at the NSLS gives the advantage of millisecond time scales with single base pair resolution. Experimental details of this technique applied to the kinetics of RNA folding can be found in previous publications (9, 10, 11, 12). X-ray radiolysis of water generates hydroxyl radicals that can be used to cleave DNA. Protein binding to DNA yields protection against hydroxyl radical attack along nucleic acid surfaces that becomes excluded from solvent. Developing protection on the nucleic acid can then be measured as a function of time in order to obtain kinetic rates of association for a protein with DNA.

Figure 1a shows the increasing protections against hydroxyl radical attack of specific DNA sequences in IHF binding site as a function of time. For this experiment, ^{32}P radiolabeled DNA was mixed with IHF and incubated for a variable amount of time in each experiment before the solution was exposed to the X-ray beam. Each lane (or vertical row) of the gel shown in Figure 1a resolves into separate bands, DNA lengths generated after hydroxyl radical attack. Protection appears simultaneously at the DNA binding sites labeled A, B, and C. Figure 1b shows the superimposed kinetic curves fit to a bi-exponential equation for each of the three binding sites. All three DNA binding sites display similar binding kinetics, indicating that they are contacted concertedly in the last step of IHF-DNA complex formation. The best fit of the data to a bi-exponential equation is characterized by an initial burst phase in binding that displays rates faster than the speed of conventional diffusion (around 10^9 M^{-1} sec^{-1}) followed by a slower step in binding corresponding to a bimolecular association rate around 10^8 M^{-1} sec^{-1}.

Figure 2 shows the IHF-DNA structure and indicates where each DNA binding site is found along the

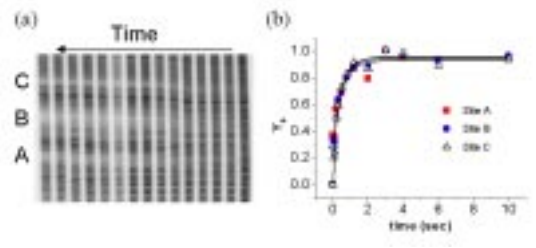

Figure 1

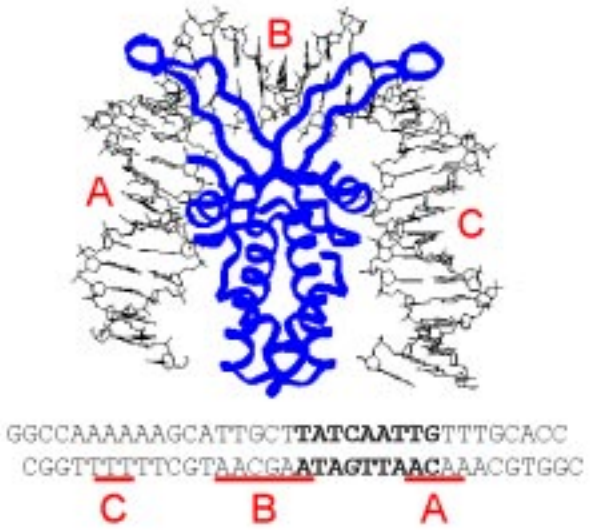

Figure 2

DNA helix. The construct used in our studies extends beyond the DNA depicted in Figure 2 and we propose that, IHF can find its binding site by sliding along the DNA in a one-dimensional diffusion search. The initial, fast step in complex formation is sensitive to salt concentration, characteristic of facilitated diffusion of proteins along the helix observed for other DNA binding proteins (13, 14). The last, slower phase is the final step in complex formation in which the sequence specific DNA contacts are formed.

Since IHF-DNA complex formation is a relatively fast process approaching the speed of diffusion, the fast time scales that can be sampled by our experimental setup at the X28C beamline, dedicated to time resolved footprinting, are essential for our kinetic studies. Additionally, the single base pair resolution provides a direct way to resolve the kinetics of protein-DNA contacts at separate sequences.

References

1. J.F. Thompson and A. Landy, "Empirical estimation of protein-induced DNA bending angles: applications to site-specific recombination complexes." Nucleic Acids Research **16**(20): 9687-9705, (1988).
2. N.L. Craig and H. A. Nash ,"*E. coli* integration host factor binds to specific sites in DNA." Cell **3**: 707-716, (1984).
3. J. Gardner and. H.A. Nash, "Role of *Escherichia coli* IHF protein in lambda site-specific recombination." Journal of Molecular Biology **191**: 181-189, (1986).
4. D.S. Hwang and A. Kornberg, "Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF." Journal of Biologicial Chemistry **267**: 23083-23086, (1992).
5. W. Xin and M. Feiss, "Function of IHF in lambda DNA packaging I. Identification of the strong binding site for Integration Host Factor and the locus for intrinsic bending in cosB." Journal of Molecular Biology **230**: 492-504, (1993).
6. J.W. Winkelman and G.W. Hatfield, "Characterization of the Integration Host Factor binding site in the ilvPG1 promoter region of the ilvGMEDA operon of *Escherichia coli*." Journal of Biologicial Chemistry **265**: 10055-10060, (1990).
7. P.A. Rice, S. Yang, K. Mizuuchi and H.A. Nash, "Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn." Cell **87**: 1295-1306, (1996).
8. H.A. Nash and C.A. Robertson, "Purification and properties of the *Escherichia coli* protein factor required for lambda integrative recombination." Journal of Biologicial Chemistry **256**: 9246-9253, (1981).
9. B. Sclavi, S.A. Woodson, M.R. Chance and M. Brenowitz, "Time-resolved Synchrotron X-ray "Footprinting", a New Approach to the Study of Nucleic Acid Structure and Function: Application to Protein-DNA Interactions and RNA folding." Journal of Molecular Biology **266**: 144-159, (1997).
10. B. Sclavi, M. Sullivan, S. A. Woodson, M.R. Chance and M. Brenowitz, "Following the Folding of RNA with Time-resolved Synchrotron X-ray "Footprinting."." Methods in Enzymology **295**: 379-402, (1998).
11. B. Sclavi, M. Sullivan, M.R. Chance, M. Brenowitz and S. A. Woodson, "RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting." Science **279**: 1940-1943, (1998).
12. C.Y. Ralston, B. Sclavi, M. Sullivan, M. Deras, S. Woodson, M.R. Chance and M. Brenowitz ,"Time-resolved Synchrotron X-ray Footprinting and its Application to RNA Folding." Methods in Enzymology **317**: 353-368, (2000).
13. R. Ruusala and D.M. Crothers, "Sliding and intermolecular transfer of the lac repressor: Kinetic perturbation of a reaction intermediate by a distant DNA sequence." Proceeding of the National Academy of the Sciences **89**: 4903-4907, (1992).
14. R.B. Winter, O.G. Berg, and P.H. von Hippel, "Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interactions: kinetic measurements and conclusions." Biochemistry **20**: 6961-6977, (1981).

Structure of Mammalian Poly(A) Polymerase in Complex with an ATP Analog

G. Martin, W. Keller[1], and S. Doublié[2]
[1]*Biozentrum, Basel, Switzerland*
[2] *University of Vermont*

Following synthesis by RNA polymerase II, eukaryotic mRNA precursors undergo a series of modifications, including capping at the 5'-end, splicing, and addition of a poly(A) tail at their 3'-end. Polyadenylation has been shown to affect virtually all aspects of mRNA metabolism, including translation initiation, mRNA stabilization and transport. In mammals, the synthesis of the poly(A) tail is catalyzed by poly(A) polymerase (PAP) with the help of the cleavage and polyadenylation specificity factor (CPSF) and poly(A) binding protein II (PABP2). As a first step towards getting a high resolution structure of the mammalian polyadenylation machinery, we have established a collaboration with Walter Keller at the Biozentrum (Basel, Switzerland) to undertake a crystallographic study of poly(A) polymerase in complex with its substrates.

Several constructs of poly(A) polymerases were tested during crystallization trials, including full-length bovine and *Caenorhabditis elegans* PAPs and deletion mutants of the bovine enzyme. We readily obtained crystals of a C-terminal deletion mutant of the bovine PAP. The C-terminal deletion has a modest effect on activity and, most importantly, this deletion mutant still binds to PABP2, and is stimulated by CPSF [1]. In complex with 3'-dATP and $MnCl_2$, the recombinant selenomethionyl PAP yielded crystals suitable for structure determination by multiwavelength anomalous diffraction (MAD). The MAD data were collected at beamline X12C at the National Synchrotron Light Source using the Brandeis B1 CCD detector. The crystal structure was determined at 2.5 Å by a combination of MAD and multiple isomorphous replacement, followed by density modification.

Poly(A) polymerases specifically incorporate ATP at the 3'-end of mRNA in a template-independent manner. These RNA polymerases belong to a large superfamily of nucleotidyl transferases containing two signature features: a conserved catalytic domain with three invariant carboxylates that are crucial for activity, and a helical turn motif involved in nucleotide binding [1]. This family also comprises mammalian DNA polymerase beta (pol β) [2] and kanamycin nucleotidyl transferase [3]. Pol β is a template dependent DNA polymerase involved in the base excision repair pathway. Until now, a structure of a template-independent polymerase such as poly(A) polymerase was lacking.

The PAP structure revealed expected and unexpected similarities to other proteins [4]. A large number of DNA and RNA polymerases share a common architecture, which has been described as a half opened right hand, composed of fingers, palm, and thumb subdomain [5]. Unlike all polymerases solved to date, PAP is a template-independent RNA polymerase, and its structure is a departure from the common polymerase fold. Interestingly, the PAP polymerase active site is located on one side of the U, and not at the confluence between the fingers and thumb (Figure 1). The hand metaphor, therefore, does not hold for PAP. For this reason, we refer to the three separate domains, from the N- to C-terminus, as the catalytic, central, and RNA binding domains.

Despite the fact that PAP does not use a nucleic acid template, incorporation of nucleotides in the growing poly(A) tail is highly specific for ATP. This poses an interesting problem: how does this template-independent RNA polymerase select for ATP? In template-dependent DNA polymerases, the shape of the active site dictates which nucleotide will be incorporated. Both the polymerase and template act in concert to assemble an active site that will select for Watson-Crick base pairs [2], [6]. In the case of PAP, the nucleotide selection proceeds in the absence of a template. Crystallographic analyses of template dependent DNA polymerases showed that there are two metals in the polymerase active site that participate in catalysis [5]. In order to identify the position of the metals in the PAP active site, we calculated an anomalous difference Fourier map (Figure 2), which revealed that there are three manganese ions in the PAP active site, instead of the two metals usually seen in DNA polymerases. The metals make key interactions that help position the nucleotide in the active site. In our crystal structure, metal A contacts N7 of the adenine base and the N6 amino group donates a hydrogen bond to Asp 167, the third aspartate of the catalytic triad (Figure 2). The adenine moiety is further stabilized in the active site of PAP via hydrophobic interactions with two conserved valines, Val 154 and Val 156.

These hydrogen bond and van der Waals contacts to the adenine base may provide an answer to the question of how PAP distinguishes adenine from the other RNA bases. The discrimination against the guanine

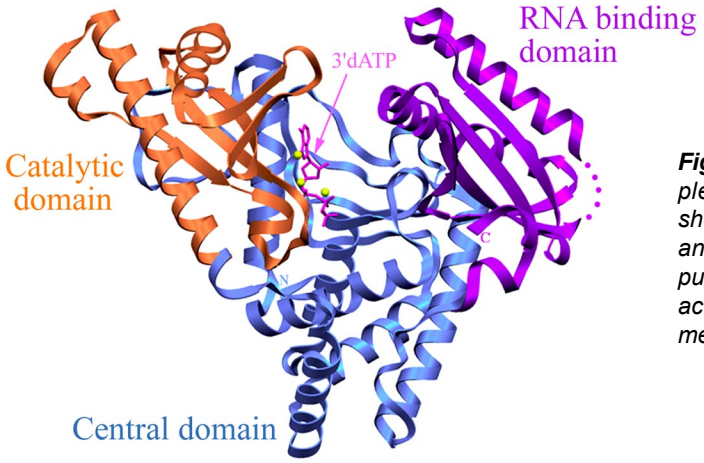

Fig. 1. *Ribbon diagram of the bovine PAP complex with 3'-dATP. The catalytic domain is shown in orange, the central domain in blue, and the C-terminal RNA-binding domain in purple. The 3'-dATP molecule bound in the active site is shown in magenta, and the three metals ions are shown as yellow spheres.*

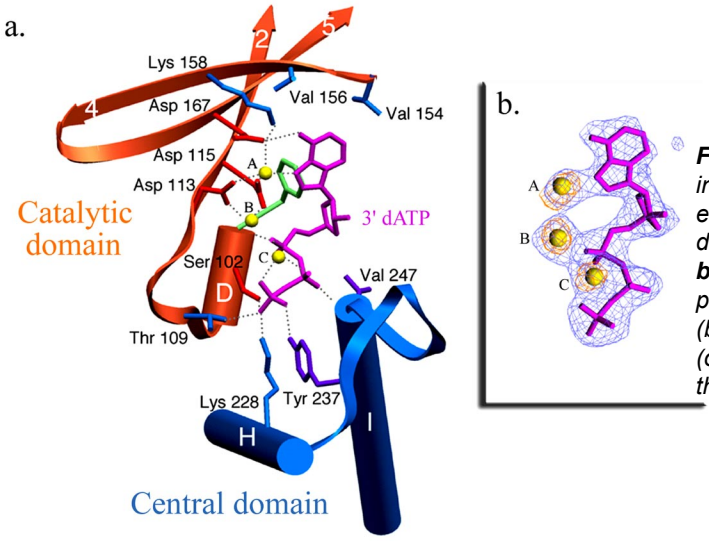

Fig. 2. a*. PAP active site showing the major interacting residues. The secondary structure elements are color-coded as in Figure 1. Hydrogen bonds are shown in dashed gray lines.* ***b.*** *The refined model of 3'-dATP is superimposed on a simulated annealing omit map (blue) and anomalous difference Fourier map (orange), contoured at 3.0σ and 4.5σ above the mean, respectively.*

base is achieved through interactions to C6 and C2. The C6 carbonyl of guanine would interact unfavorably with Asp 167. Furthermore, Val 154 makes a van der Waals contact to C2, which would preclude binding of any base with a substituent at that position, therefore providing a structural basis for discrimination against GTP incorporation.

This study of bovine PAP complexed to 3'-dATP provides the first structure of any component of the mammalian polyadenylation machinery and the first example of a template-independent polymerase. This crystal structure will guide additional studies of PAP complexed to poly(A) RNA and proteins of the polyadenylation machinery that are known to interact with the polymerase.

References:

[1] G. Martin and W. Keller EMBO J. **15**, 2593, 1996.
[2] M.R. Sawaya, R. Prasad, S.H. Wilson, J. Kraut and H. Pelletier Biochemistry **36**, 11205, 1997.
[3] L.C. Pedersen, M.M. Benning and H.M. Holden Biochemistry **34**, 13305, 1995.
[4] Martin G., W. Keller and S. Doublié EMBO J. **19**, 4193, 2000.
[5] T. A. Steitz *J. Biol. Chem.* **274**, 17395, 1999.
[6] S. Doublié, S. Tabor, A.M. Long, C.C. Richardson and T. Ellenberger Nature **391**, 251, 1998.

Protein Crystal Derivatization by Short Cryo-Soaking with Halides

Z. Dauter[1,2], M. Dauter[1,2] and K. R. Rajashankar[2,3]
[1]SAIC/National Cancer Institute, Frederick, MD 21702, USA
[2]Brookhaven National Laboratory, Bldg. 725A-X9, Upton, NY 11973, USA
[3]The Rockefeller University, 1230 York Avenue, New York, N.Y. 10021, USA

The phase problem remains an unsolved puzzle in protein crystallography. Macromolecular crystallographers have been relying on heavy atoms in protein molecules to obtain the phase information. In a few cases, heavy atoms like Cu, Fe or Zn may be inherently present in the protein. However, very frequently, heavy atoms have to be incorporated into protein molecules by chemical methods. One of the classical methods of introducing heavier metals like Hg, Au, Pt, etc. is to soak the protein crystals in these heavy metal salt solutions. They bind covalently to certain functional groups, provided that the protein molecule offers the required chemical environment for heavy metals. When the required chemical environment is not present, heavy metals fail to derivatize protein molecules. A recent and more popular method of derivatization is to introduce selenomethionine residues by recombinant techniques. However, this method does not suit certain expression systems. In addition, the number of methionine residues present in the protein molecule can be a limiting factor.

In the present work, we demonstrate an easy and rapid way of introducing anomalously scattering halide ions into protein crystals, thereby providing a fast and simple route to derivatize protein molecules. In brief, bromide or iodide anions were introduced into the cryoprotecting solution and protein crystals were soaked in this cryosolution for a short duration prior to cryogenic freezing. The halide ions rapidly diffuse into protein crystals and occupy ordered sites on the surface of the protein molecules. So introduced, halide sites can be successfully utilized for solving the phase problem. It should be noted that one can perform MAD experiments on bromide soaked crystals, as bromide has a K-absorption edge at 0.92Å (13474 eV), which is often accessible at MAD capable synchrotron beamlines. In contrast, absorption edges of iodide are far from accessible, however, iodide soaked crystals are ideal for the SAD or SIRAS approach of phasing, as iodide offers ~6.9 anomalous electrons at 1.54Å (8050 eV).

This method of derivatization was tested on four proteins with known structures. These four proteins are: lysozyme (McPherson, 1982), Rnase A (Kartha et. al. 1967), subtilisin (Betzel et. al. 1988), and xylanase (Viswamitra et. al. 1993). These were crystallized according to the methods described in the literature (Table 1). As is evident from Table 1, the test proteins were crystallized from various concentrations of protein, diverse precipitating agents and at different pH. Cryoprotecting solutions were prepared by including certain amounts of glycerol in the mother liquor (except for RNase A, which was crystallized from MPD solution, a good cryoprotectant). A key point, which is specific to this method, is that cryoprotecting solutions also contained high concentrations of either NaBr or KI. The crystals were transferred to the cryoprotecting solution for a short time, between 15 and 45 sec and frozen in cryostream. A 1 M concentration of NaBr did not affect the quality and diffracting power of the crystals, but 1M KI negatively influenced crystals in certain cases where the KI concentration was reduced to lower values (Table 1). X-ray diffraction data were collected at the National Synchrotron Light Source, Brookhaven National Lab., beamline X9B, using either ADSC Quantum4 or MAR 165mm CCD detectors. The rotation range was selected to yield complete anomalous data, but no inverse beam technique was used. Though crystals of all four test proteins diffract to atomic resolution, the exposure times were short and diffraction data were recorded to a maximum resolution ranging from 1.4 to 1.7 Å (Table 2). The diffraction images were processed with HKL2000 suite (Otwinowski & Minor, 1997). The results are summarized in Table 2.

The halide sub-structures were determined by using the direct method program SHELXS (Sheldrick, 1990) or SnB (Miller et al. 1994), utilizing only the anomalous differences. In all cases, clearly indicated correct solutions (within the equivalent sets of possible origin shifts or enantiomers) were obtained. All phasing calculations were performed with the program MLPHARE (Otwinowski, 1991). For Br-MAD data of Lysozyme, Subtilisin and Xylanase, the data collected at inflection point wavelength were accepted as native and two other wavelengths as derivatives. The single wavelength anomalous dispersion (SAD) phasing was applied to Rnase A soaked in bromide. For iodine derivatives, the SIRAS approach was adopted utilizing the available native data. The selection of peaks from SHELXS for use for phasing in MLPHARE was somewhat arbitrary, since their heights diminish gradually

from the maximum value without clear contrast between correct and false sites. About eight to ten highest peaks were accepted from the best trial and additional sites were identified from the anomalous-difference Fourier synthesis. The phases obtained from MLPHARE were subjected to density modification using DM (Cowtan, 1999). The values of figure of merit and correlation coefficient with the Fobs map for all protein residues are given in Table 2. An example of the resulting DM-phased electron density is shown in Figure 1. In all test cases clearly interpretable maps were obtained which were further improved and autotraced by using the program wARP (Lamzin & Wilson, 1997).

Halides are not coordination specific

Inspection of the halide sites in these test structures showed some interesting results and also explained why this method works. The first and foremost point is that halide ions, unlike classical heavy atoms, are not specific for a particular type of chemical coordination. Since halide ions are polarizable they adopt themselves to the chemical environment available around the protein surface. Based on this reasoning, one can speculate halides to derivatize virtually any protein or biomolecule in general. Of course, locating these halide sites and utilizing them for phasing requires accurate and good quality data, as in the case of any good-practiced crystallographic undertaking. Figure 2 shows a few examples of various halide-protein interactions. Most of them have H-bond contacts with hydrogen donor groups of protein or water on one side, and hydrophobic interactions in non-polar niches at the protein surface on the other. Those halide ions, which form ion pairs with positively charged arginine or lysine side chains, have the highest occupancies. Since arginines and lysines have high pK_{α} values, they remain positively charged up to a relatively high pH, and are able to strongly bind to halide anions in a wide pH range of crystallization conditions.

Halides are position specific

Halides are specific for the sites they bind to in a given protein. What this means is that if ten crystals of a protein are soaked in halides, all crystals will have the same halide sites. This is best demonstrated by the structure of the human acyl protein thioesterase I, one of the unknown structures solved by this novel method (Devedjiew et al. 2000). Crystals of this protein contained two monomers of 28 KD each in the asymmetric unit. It was phased using 22 bromides employing the SAD technique. Out of these 22 bromides, 18 of them adhere to the non-crystallographic 2-fold present in the dimer (Figure 3). Moreover, the way in which they interact with the protein are comparable in individual monomers. This clearly points to the position specificity of bromides.

Bromides and Iodides occupy the same sites

Bromide and iodide anions, in general, occupy the correspondingly same ordered solvent sites around the protein surface, within the ordered solvent shell. This is illustrated in Figure 4a. One can see that most of the peaks in the anomalous-difference Fourier map of bromide soaked Xylanase crystals have corresponding

Table 1

Crystallization conditions for the four test proteins

Protein	Rnase A	Lysozyme	Subtilisin	Xylanase
Size (a.a)	124	129	269	301
Space group	$P2_1$	$P4_32_12$	$P2_12_12_1$	$P2_1$
Crystallization conditions	50% MPD 100 mM NaAc pH 5.4	1 M NaCl 100 mM NaAc pH 4.7	12 % PEG 4K 100 mM citrate 1 M NaCl 10 mM $CaCl_2$ pH 6.0	10 % $AmSO_4$ 100 mM Tris/HCl pH 7.4
Cryosolution	50 % MPD 100 mM NaAc 1 M NaBr or 0.35 M KI	100 mM NaAc 30 % glycerol 1 M NaBr or 0.5 M NaCl + 0.5 M KI	12 % PEG 4K 10 mM citrate 10 mM $CaCl_2$ 25 % glycerol 1 M NaBr or 1 M KI	10 % $AmSO_4$ 100 mM Tris/HCl 25 % glycerol 1 M NaBr or 1 M KI

Table 2
X-ray data collection and Phasing details

Protein	Rnase A		Lysozyme		Subtilisin		Xylanase	
Soak	Br	I	Br*	I	Br*	I	Br*	I
Resolution (Å)	1.50	1.62	1.70	1.62	1.40	1.60	1.50	1.60
Rmerge (%)	3.5	5.9	3.8	5.0	5.1	4.4	2.4	4.4
Completeness (%)	99.9	95.1	100.0	99.5	99.9	93.6	88.5	94.4
$I/\sigma(I)$	28.7	15.7	30.4	20.0	21.2	26.5	30.0	25.6
Anomalous Multiplicity	3.7	3.2	3.8	3.2	3.1	3.3	1.9	3.0
Phasing method	MAD	SIRAS	SAD	SIRAS	MAD	SIRAS	MAD	SIRAS
No. of sites used for phasing	7	11	12	9	7	8	16	14
FOM after DM	0.87	0.77	0.73	0.74	0.85	0.76	0.78	0.75
Corr. Coeff. with Fo map	0.89	0.73	0.68	0.63	0.80	0.68	0.65	0.69
No. of sites > 10σ (ΔF^{anom}, φ_{calc}-90°) map	8	8	9	5	17	16	18	11

* Statistics for the remote wavelength is given. Data for other wavelengths are very similar.

peaks in the Iodine soaked crystals. However, iodide seems to be more penetrating than bromine in specific cases. Figure 4b demonstrates an interior iodide in xylanase. In contrast, bromide soaked xylanase does not show any peak at the corresponding position. This is in agreement with our earlier observation that in certain cases the concentration of iodide had to be reduced to maintain diffraction quality if the crystals.

Halide sites are shared with waters

Halide ions partially replace certain water molecules at the surface of the protein. No correlation could be observed between refined B factors of water oxygen atoms and the peak heights at the corresponding positions in the anomalous difference Fourier syntheses. This suggested that halide ion sites are partially occupied, shared between Br^- or I^- ions and water molecules. Indeed, inspection of the native and halide soaked electron density maps show halides at water sites in most cases (Figure 5). This was also evident during halide substructure determination by direct methods - the peak heights in the E-maps ranged from a maximum, decreasing gradually to lower values, in contrast to the usual situation in small molecule crystallography or SeMet data, wherein right peaks clearly stand out from noise.

Number of halide sites depends upon protein size and halide concentration

The number of halide sites seems to depend mainly upon two factors, the protein size and the halide concentration. In Table 2, in the case of bromide soaks, the number of anomalous-difference Fourier peaks greater than 10s is proportional to the protein size. However, in the case of iodide soaks, this is not the situation. Though Lysozyme and Rnase A are similar sized proteins, Lysozyme has more peaks than Rnase A and so is the case with Subtilisin and Xylanase. This

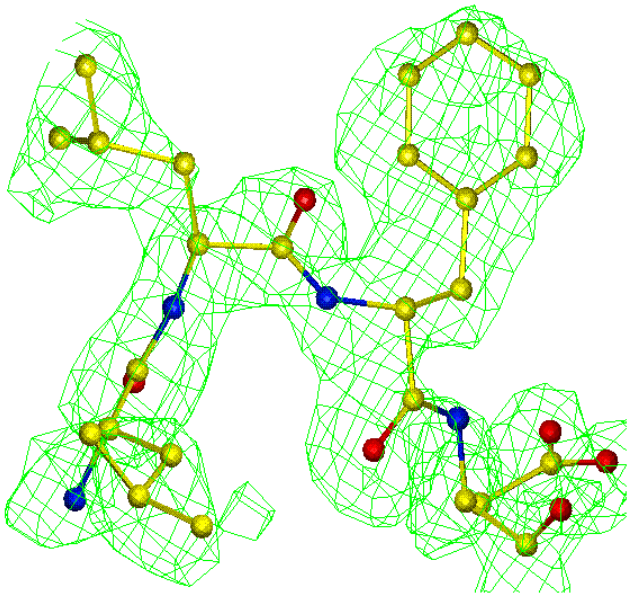

Fig. 1 Representative portion of the SIRAS experimental electron density map after density modification by DM at 1σ level for iodide soaked xylanase.

can be rationalized if we take in to account the soaking concentration of KI, which was reduced in the case of RnaseA and Xylanase. In addition, our preliminary analysis shows that the number of halide sites does not depend much on the soaking time – 30 secs is as good as 2 mins.

Application to unknown structure

To the best of our knowledge, around ten new structures have been solved using the halide soak method so far. The GAF domain (36KD, Ho et al. 2000), Human acyl Protein Thioesterase I (56KD, Devedjiew *et al*. 2000), β-defensin (13KD, Hoover *et al*. 2000) and pepstatin-insensitive carboxyl proteinase (44KD, Dauter *et al*. 2000) are some of the published results, and there are a few more which are yet to be published. In some of these cases, the current method made a big difference for the structure solution, as all efforts to derivatize these specific proteins by other means were futile. These results demonstrate the general feasibility of this method to solve unknown structures.

Bromide or iodide ions, when present in the cryoprotecting solution at high concentrations, rapidly diffuse into protein crystals. This diffusion is so fast that even about a ½ minute soaking will result in ordered halide sites on the protein surface. The anomalous signal from these halide sites can be used to solve the phase problem in protein crystallography. This method has been tested for four different proteins and utilized to solve several unknown protein structures.

These halide ions are not specific for any chemical coordination. Hence, it can be argued that halides can derivatize virtually any protein by this method.

The speed and ease of incorporating halide ions into protein crystals by the current method makes it particularly applicable for high throughput crystallographic projects. It can be used as an alternative way to the popular Se-methionine based MAD and classical heavy atom MAD/MIR approach to phase protein structures. The generality of derivatization finds applications for stagnant MR problems wherein some phases can be of help.

Acknowledgments

We thank Boris Strokopytov for providing crystals of RNase A, Prof. Viswamitra and Prof. Ramakumar for crystals of xylanase and Mike Sullivan for help with the beamline and detector setup.

References

A. McPherson, *The preparation and analysis of protein crystals*. Wiley, New York (1982).

G. Kartha, J. Bello and D. Harker, Tertiary structure of ribonuclease, *Nature*, **213**, *862*, 1967.

C. Betzel, Z. Dauter, M. Dauter, M. Ingelman, G. Papendorf, K.S. Wilson, and S. Branner, *J. Mol. Biol.*, **204**, *803*, 1988.

M.A. Viswamitra, P. Bhanumoorthy, S. Ramakumar, M.V. Manjula, P.J. Vithayathil, S.K. Murthy, and A.P. Naren, Crystallization and preliminary X-ray diffraction analysis of crystals of Thermoascus aurantiacus xylanase. *J. Mol. Biol.*, **232**, *987*, 1993.

Z. Otwinowski, and W. Minor, Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol.*, **276**, *307*, 1997.

G.M. Sheldrick, *Acta Cryst.*, **A46**, *46*, 1990.

R. Miller, S.M. Gallo, H.G. Khalak and C.M. Weeks, SnB – Crystal structure determination via Shake-and-Bake, J. Appl. Cryst., **27**, *613*, 1994.

Z. Otwinowski, *Proceedings of the CCP4 Study Weekend.* (eds. Wolf, W., Evans, P.R. and Leslie, A.G.W.) (SERC Daresbury Laboratory, Warrington, UK), *80*, 1991.

K.D. Cowtan and K.Y.J Zhang Density modification for macromolecular phase improvement, *Prog. Biophys. Mol. Biol.*, **72**, *245*, 1999.

V.S. Lamzin and K.S. Wilson, Automated refinement for protein crystallography, *Meth. Enzymol.*, **277**, *269*, 1997.

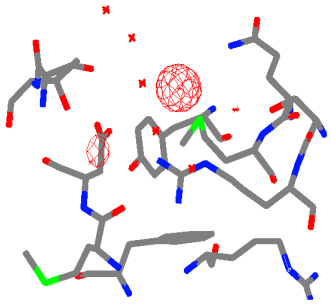

Figure 2a

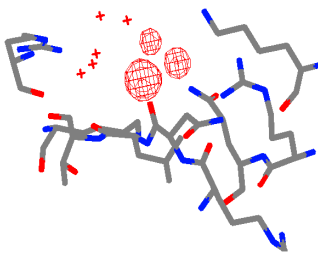

Figure 2b

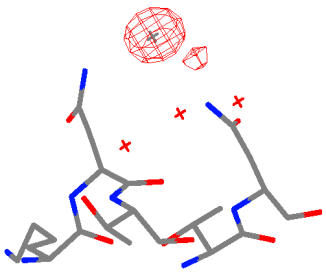

Figure 2c

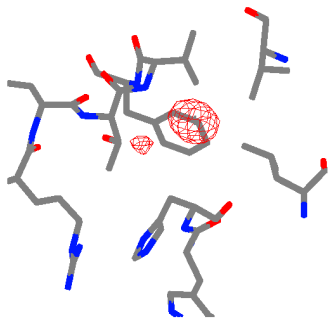

Figure 2d

Fig. 2 The environment of halides in protein structures. The anomalous difference density is also shown at 5σ. (a, b) Bromide ions in Rnase A near arginine side chains (c) Iodide ion coordinated by two asparagine NH_2 groups at the surface of xylanase and (d) iodide ion located in a hydrophobic niche and hydrogen bonded to a main chain amide of xylanase.

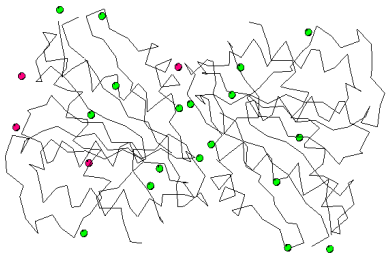

Fig. 3 The backbone of the human acyl protein thioesterase I along with the bromide sites used for phasing. The bromides colored green follow the non-crystallographic 2-fold symmetry of the protein molecule and pink ones don't. This is a consequence of position specificity of bromide sites

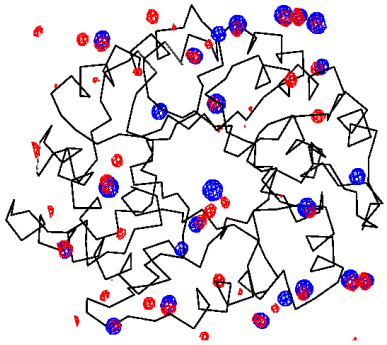

Figure 4a

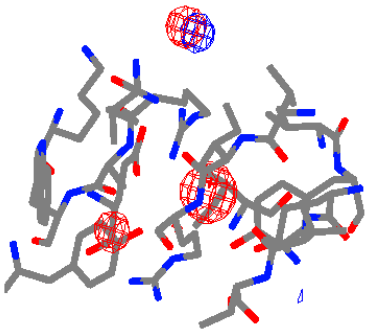

Figure 4b

Fig. 4 Anomalous difference Fourier synthesis (ΔF^{anom}, φ_{calc}-90°) for Xylanase showing bromide (red) and iodide (blue) sites, contoured at 5σ. (a) Notice that most of the bromide and iodides occupy the same sites. (b) An interior iodide ion in iodide soaked xylanase crystal. The corresponding site is not occupied in the bromide soaked crystal. The red peak on the left-hand bottom side corresponds to sulfur from a methionine residue.

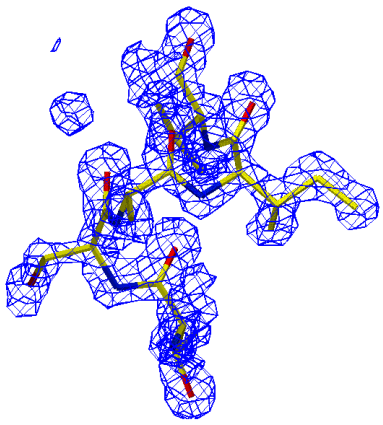

Figure 5a

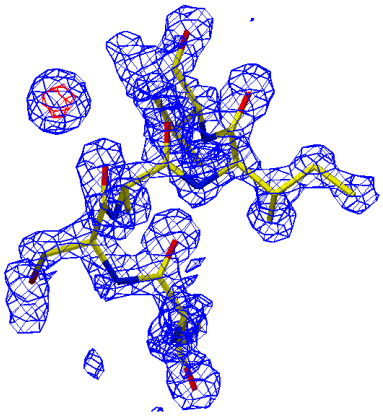

Figure 5a

Fig. 5 wARP refined map of (a) native xylanase (b) bromine soaked xylanase. Maps contoured at 1.2σ (blue) and 5σ (red). The red peak in (b) is a bromine ion occupying the site of a water molecule in (a).

Y. Devedjiew, Z. Dauter, S.R. Kunzetsov, T.L.Z. Jones and Z.S. Derewenda, Crystal structure of the Human Acyl Protein Thioesterase I from Single X-ray Data Set to 1.5Å, Structure Fold Des., **8**, *1137*, 2000.
Y.J. Ho, L.M. Burden and J.H. Hurley. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor, EMBO J. **19**, *5288*, 2000.
D.M. Hoover, K.R. Rajashankar, R. Blumenthal, A. Puri, J.J. Oppenheim, O. Chertov and J. Lubkowski, The structure of Human β-Defensin-2 Shows Evidence of Higher-Order Oligomerization, J. Biol. Chem. **275**, *32911*, 2000.
Z. Dauter, M. Li and A. Wlodawer, Practical experience with halides for phasing macromolecular structures: a powerful tool for structural genomics, Acta Cryst. **D57**, 2001 (accepted).

Crystal Structure of Lyme Disease Antigen, OspC from HB19 Strain

D. Kumaran, S. Eswaramoorthy, J. Dunn and S. Swaminathan
Biology Department, Brookhaven National Laboratory

Lyme disease is a tick-borne infectious disease caused by the spirochete *Borrelia burgdorferi*. The associated clinical manifestations are variable early during infection, involving skin, joints, heart and the nervous system, and may become chronic at later stages. The causative agent *Borrelia burgdorferi* induces a strong humoral response against endoflagellar protein, p41, and a protein constituent of the protoplasmic cylinder, p93, both of which are enveloped within the outer membrane, and some outer surface lipoproteins (Osps), which are major membrane components.

It has been shown that when a tick first becomes infected, the Lyme spirochetes settle in its midgut and make a surface protein called OspA. However, when an infected tick subsequently attaches to a warm-blooded mammal and begins feeding on blood, the spirochete stops producing OspA and makes the OspC surface protein instead. The switch is in part due to the change in temperature; OspC is induced at 32°C to 37°C, but not at 24°C, and this up regulation is at the transcriptional and translational levels. Evidence suggests co-regulation of these two genes at the mRNA level. Clearly, to survive in both hosts, spirochetes have evolved mechanisms for sensing the different host environments and responding accordingly. Experimental OspA and OspC vaccines have limited utility since they are usually only effective against challenge by the same strain and not by heterologous strains.

The *osp*C gene is located on a 27-kb circular plasmid and encodes a lipoprotein of 22 to 23 kDa. OspC proteins are highly polymorphic and this variability extends even to a single *Borrelia* population. For example, alleles of OspC collected from a single site on Shelter Island, New York, could be clustered into 19 major groups or types (A-S) based on sequence homology[1]. Sequence variation within a major group is less than 1% but about 15% across the major groups. Variation within a local population is comparable to the variation of similar size samples collected from the entire species. This variability has consequences in the development of OspC-based seradiagnostic antigens and anti-OspC vaccines. Of the 19 major groups, only four (A, B, I and K) contain invasive clones and cause infections at skin and extracutaneous sites while the others are non-human pathogens or they infect only the skin[2]. However, the biological function of OspC is not known nor is the relationship between OspC type and pathogenicity understood. In order to develop an effective OspC-based vaccine, it is important to know representative three-dimensional structures of at least a few OspCs, especially across the major groups. Here we report the crystal structure of OspC from strain HB19 (aa 38-201) which is a member of invasive group I.

Se-met OspC crystals were grown at 293K using the sitting drop vapor diffusion method[3]. Diffraction data extending to 2.8 Å were collected from a frozen crystal using the Brandeis detector on beamline X12C at the NSLS. Three-wavelength MAD data were collected around the selenium absorption edge. The data were processed using DENZO and SCALEPACK. Subsequently, single wavelength higher resolution data extending to 1.8 Å were collected with a better quality crystal at the X25 beamline at the NSLS. A total of 218,740 measurements to 1.8 Å were reduced to 53,350 unique reflections with the R_{merge} of 0.087 and with a completeness of 85.3%. This high-resolution data were used in the later stages of refinement.

There are two dimers per asymmetric unit with two selenomethionines per monomer. The selenium positions were obtained from the Patterson and difference

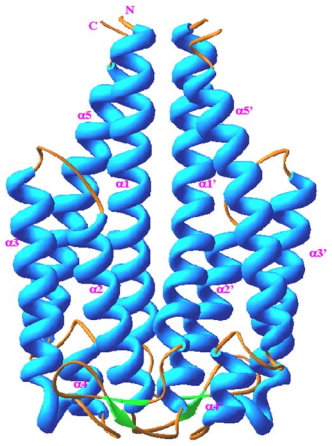

***Figure 1**. RIBBONS representation of HB19 dimer. The close proximity of the two inner-most helices indicates that it is a tight dimer.*

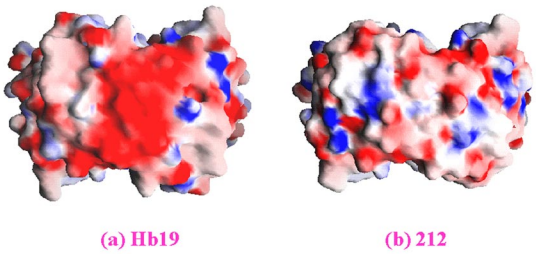

***Figure 2**. Electrostatic surface potential of various OspCs. Electrostatic surface potential of (a) HB19 and (b) 212 as viewed from the top of the molecule. The negative and positive potentials are shown in red and blue. The HB19 and 212 spirochetes are classified into invasive and non-invasive groups, respectively. The electrostatic surface potential at the top of the OspC dimers from spirochetes in the invasive group is highly negative and clearly different from those of the non-invasive group.*

Fourier maps using the PHASES program package. The dimers are formed *via* a non-crystallographic two-fold symmetry and are related by a pseudo translation. A total of 8 selenium atoms were input to the SHARP program to refine the phases. The resulting phases were further improved by NCS averaging in DM in two steps, averaging the non-crystallographic two-fold related monomers first and then the dimers related by pseudo translational symmetry. This resulted in a relatively unambiguous electron density map for tracing the backbone and identifying most of the residues in the helical regions. The model building was completed using the two selenomethionines as markers.

The model was subjected to rigid-body refinement using the program CNS with the high-resolution data set followed by the slowcool annealing procedures. A composite omit map clearly showed the electron density for zinc ions and including these zinc ions further reduced the R-value. The final model contains 4 monomers of 162 residues each; six zinc atoms and 632 water molecules. Two N-terminal residues were not included in the model for all the monomers because of poor electron density. Final R and free-R are 0.20 and 0.23, respectively, for 47,286 reflections between 50 and 1.8 Å resolution.

The structure of the OspC monomer resembles a classical up and down α-helical bundle. This is in contrast to the structure of OspA which has an unusual, elongated fold composed of 21 anti-parallel β-strands followed by a single small α–helix[4]. Except for helices α4 and α5 (Figure 1), all helices are anti-parallel. Two monomers form a 10-helical bundle dimer through a non-crystallographic two-fold axis almost parallel to helices α1 and α1'. The dimeric interaction at the interface is almost hydrophobic. The buried surface area between the dimers is 4430 Å^2, which is about 30% of the total surface area of the dimer (14374 Å^2) suggesting that the dimer interaction is very strong, leading us to believe that the biologically functional molecule is a dimer. The dimeric interaction is mostly between helices α1 and α1', which are parallel. Additional interactions occur between helices α1 and α2' and between the loop connecting α1 and α2, and the helix α3'. The shape of the dimer resembles elongated kidneys with the approximate overall dimension of 62x40x27 Å^3. There are three zinc ions per dimer, which take part in packing interactions. Interestingly, three different coordinations, *viz*., tetra, penta and hexa coordinations of zinc are observed. The presence of zinc atoms in the crystal is probably due to zinc sulfate, which was used as an additive during crystallization.

Dali database search indicates that OspC may be very similar to the periplasmic domain of the Salmonella aspartate receptor (AR), which is also a dimer[5]. The superposition of a monomer of AR on a monomer of HB19 gave an rmsd of 3.0 Å for 122 aligned Cα atoms. Although the three-dimensional alignment seems to be very good, the overall sequence homology between OspC and the periplasmic AR domain is low, with only 13% identity; thus the similarity between the two proteins would have been missed based on sequence alone.

Several studies have shown that *B. burgdorferi* have a predilection for collagenous tissue and can interact with fibronectin and cellular collagens. The spirochetes can bind to a number of different cell types, including fibroblasts. *B. burgdorferi* can bind to a novel circulating fibroblast-like cell called the peripheral blood

fibrocyte that expresses collagen types I and III, as well as fibronectin, in a process that does not require OspA or OspB. Interestingly, the electrostatic potential at the surface that projects away from the membrane is highly negative (Figure 2a). This region is lined by carbonyl oxygens, aspartate and glutamate residues. Two deep 50 Å cavities were also noted in this area. If OspC is indeed a binding protein, its binding interface may involve this region. Also, if it were to bind to fibronectin or a similar molecule, the binding surface in the ligand would have to be positively charged.

The ultimate aim of this structure determination is to obtain a representative structure of OspC, to use this information to model other OspCs and to understand the differences between various OspCs. Therefore, we modeled 18 other OspCs, one representative from each of the other 18 groups. All are predicted to fold similarly. Multiple sequence alignment showed that the variable regions in these sequences are mainly in the loop regions of the molecule. Presumably, these variable loop regions constitute major antigenic sites that serve to differentiate amongst OspC antigens. The information will be useful for designing chimeric OspC molecules, which might be effective against more than one OspC antigen.

Only limited sub-populations of *B. burgdorferi* are responsible for invasive human disease and OspC type can be used as means to distinguish invasive and non-invasive spirochetes. Certain types of OspC never cause human disease while some others cause local infection at the primary site skin, but do not cause any systematic disease. Only four types (A, B, I and K) are responsible for systematic disease and are found in the secondary sites and hence are termed invasive clones. When we compared the electrostatic potential of HB19, a type I molecule, with similar models for the other OspCs, a striking distinction between OspCs associated with invasive and non-invasive spirochetes has been observed. The electrostatic potential surface viewed from the farthest point from the surface of the membrane is shown in Figure 2. While the surface potential is highly negative for OspCs from invasive strains, it is not so for those present on non-invasive strains. If OspC is involved in targeting the infecting spirochetes to specific host molecules, then this difference may provide a potential explanation of why some types of OspC are found associated with invasive clones.

The three-dimensional structure of Outer surface protein C (OspC), a major antigen of the Lyme disease, has been determined to 1.8 Å resolution. The crystal structure reveals a dimeric association of OspC with each monomer showing a five helical bundle. This is in contrast to the structure of OspA, a major surface antigen when spirochetes are residing in the midgut of unfed ticks, which is mostly β sheet. The surface of OspC that would project away from the spirochete's membrane has a region of strong negative electrostatic potential, which may be involved in binding to positively charged host ligands. This feature seems to be present only on OspCs from strains known to cause invasive human disease.

Acknowledgments

We thank B. Lade for technical assistance.

References

1. I.N Wang et al, Genetic diversity of ospC in a local population of *Borellia burgdorferi* sensu stricto. Genetics **151**, 15-30, 1999.
2. G. Seinost et al, Four clones of *Borellia burgdorferi* sensu stricto cause invasive infection in humans. Infect. Immun. **67**, 3518-3524,1999.
3. D. Kumaran, S. Eswaramoorthy, J.J. Dunn and S. Swaminathan. Crystallization and preliminary X-ray analysis of *Borellia burgdorferi* Outer surface protein C (OspC). Acta Cryst. D (in press).
4. H. Li, J.J. Dunn, B.J. Luft and C.L. Lawson. Crystal structure of Lyme disease antigen outer surface protein A complexed with Fab. Proc. Natl. Acad. Sci. USA **94**, 3584-3589, 1997.
5. I.J. Yeh, et al. High-resolution structures of the ligand-binding domain of the wild-type bacterial aspartate receptor. J. Mol. Biol. **262**, 186-201, 1996.